

LSTM-MDNz: Estimating Quasar Photometric Redshifts with an LSTM-Augmented Mixture Density Network

JIANZHEN CHEN ,¹ ZHIJIAN LUO ,¹ LIPING FU ,¹ ZHU CHEN ,¹ HUBING XIAO ,¹ SHAOHUA ZHANG ,¹ AND CHENGGANG SHU¹

¹*Shanghai Key Lab for Astrophysics, Shanghai Normal University, Shanghai 200234, People’s Republic of China*

ABSTRACT

Quasar photometric redshifts are essential for studying cosmology and large-scale structures. However, their complex spectral energy distributions cause significant redshift–color degeneracy, limiting the accuracy of traditional methods. To overcome this, we introduce LSTM-MDNz, a novel end-to-end deep learning model combining long short-term memory networks (LSTM) with mixture density networks (MDN). The model directly uses multi-band photometric fluxes and associated errors as wavelength-ordered sequential inputs, eliminating the need for manual feature engineering while enabling simultaneous point estimation and probability distribution function (PDF) prediction of quasar redshifts. We integrate data from four major sky surveys—SDSS, DESI-LS, WISE, and GALEX—to assemble a sample of over 550,000 spectroscopically confirmed quasars ($0 \leq z_{\text{spec}} \leq 5$) across 14 ultraviolet to infrared bands for model training and testing. Experimental results show that using all 14 bands yields optimal performance, with a normalized median absolute deviation (σ_{NMAD}) of 0.037 and an outlier rate (f_{out}) of 3.5% on the test set. These values represent reductions of 29% and 56%, respectively, compared to the commonly adopted SDSS+WISE band set. Probability integral transform (PIT) and continuous ranked probability score (CRPS) analyses confirm that the predicted PDFs align closely with the true redshift distribution. Band-ablation experiments further highlight the essential role of ultraviolet and infrared data in alleviating color degeneracy and reducing systematic bias. This study demonstrates the effectiveness of multi-band fusion in improving quasar photo-z accuracy and offers a ready-to-use estimation framework for future surveys like LSST, CSST, and Euclid.

Keywords: Astrostatistics (1882) — Astronomy data analysis (1858) — Photometry (1234) — Sky surveys (1464) — Redshift surveys (1378) — Computational astronomy (293)

1. INTRODUCTION

Quasars, the most energetically luminous active galactic nuclei in the universe, serve as vital probes for investigating the large-scale structure of the universe, galaxy evolution, and the growth of supermassive black holes (M. Rauch 1998; P. F. Hopkins & M. Elvis 2010; J. Kormendy & L. C. Ho 2013). They also play a crucial role in elucidating the nature of dark energy and understanding the epoch of cosmic reionization (X. Fan et al. 2006; D. H. Weinberg et al. 2013; Y. Matsuoka et al. 2016). The advent of next-generation large-scale photometric surveys, such as the Legacy Survey of Space and Time (LSST), the Chinese Space Station Survey Telescope (CSST), and the Euclid space telescope, is expected to discover and identify tens of millions of quasars (Ž. Ivezić et al. 2019; Y. Gong et al. 2019; Euclid Collaboration et al. 2020a). However, obtaining precise spectroscopic redshifts (z_{spec}) for such a vast sample is observationally expensive and time-consuming, making spectroscopic confirmation impractical for the majority of sources. Consequently, efficient estimation of quasar photometric redshifts (z_{phot}) using readily available multi-band photometric data has become an indispensable approach for statistical studies of large samples (G. T. Richards et al. 2001; M. A. Weinstein et al. 2004).

Photometric redshift estimation refers to the technique of inferring cosmological redshifts from multi-band photometric measurements (e.g., magnitudes and colors) without requiring spectroscopic observations. This method enables

the rapid acquisition of three-dimensional spatial distributions for large samples, expanding the scale of observable universe studies by orders of magnitude compared to spectroscopic samples, thereby enhancing the statistical power of cosmological measurements (H. Hildebrandt et al. 2010). Nearly all current and upcoming large-scale survey projects—including the Sloan Digital Sky Survey (SDSS; M. Fukugita et al. 1996; D. G. York et al. 2000), the Dark Energy Survey (DES; Dark Energy Survey Collaboration et al. 2016; T. M. C. Abbott et al. 2021), the Kilo-Degree Survey (KiDS; J. T. A. de Jong et al. 2013), LSST (P. A. LSST Science Collaboration: Abell et al. 2009; Ž. Ivezić et al. 2019), the Euclid Space Telescope (R. Laureijs et al. 2011), the Wide Field Infrared Survey Telescope or Nancy Grace Roman Space Telescope (WFIRST; J. Green et al. 2012; D. Spergel et al. 2015; R. Akeson et al. 2019), and the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; H. Aihara et al. 2018)—rely heavily on photometric redshifts to achieve their core scientific objectives.

Current photometric redshift estimation methods are broadly categorized into two types: template-fitting methods and machine learning methods. Template-fitting techniques such as HyperZ (M. Bolzonella et al. 2011), BPZ (N. Benítez 2000), EAZY (G. B. Brammer et al. 2008), and Le Phare (S. Arnouts et al. 1999) derive redshifts by fitting photometric data with spectral energy distribution (SED) templates (K. M. Lanzetta et al. 1996; A. Fernández-Soto et al. 1999; M. Bolzonella et al. 2000; Z. Luo et al. 2024a). Although this approach has a clear theoretical foundation, its accuracy heavily depends on the completeness and accuracy of the template library. In contrast, machine learning (or training-set) methods learn the complex mapping between photometric features and redshifts from large spectroscopic training samples, employing a wide variety of models. Commonly used techniques include artificial neural networks (ANN; A. E. Firth et al. 2003; A. A. Collister & O. Lahav 2004; I. Sadeh et al. 2016), support vector machines (SVM; Y. Wadadekar 2005), self-organizing maps (M. J. Way & C. D. Klose 2012; J. E. Geach 2012), Gaussian process regression (M. J. Way & A. N. Srivastava 2006), k-nearest neighbors (kNN; N. M. Ball et al. 2007), boosted decision trees (D. W. Gerdes et al. 2010), random forests (RF; M. Carrasco Kind & R. J. Brunner 2013; M. M. Rau et al. 2015; S. Mucesh et al. 2021; J. Lu et al. 2024), and sparse Gaussian frameworks (I. A. Almosallam et al. 2016). More recently, deep learning approaches such as multi-layer perceptrons (MLP; X. Zhou et al. 2021), convolutional neural networks (CNN; B. Hoyle et al. 2015; A. D’Isanto & K. L. Polsterer 2018; J. Pasquet et al. 2019), and Bayesian neural networks (BNN; X. Zhou et al. 2022) have also been widely applied in photo-z estimation. When sufficient training samples are available, machine learning methods often achieve higher accuracy.

In recent years, with the growing volume of astronomical data and advances in computational capabilities, machine learning methods have been widely applied to photometric redshift estimation, delivering impressive results across various surveys. For instance, in galaxy photometric redshifts, S. Mucesh et al. (2021) used DES data and the COSMOS2015 catalog to show that random forest algorithms yield robust redshift estimates even with limited photometric bands. The Euclid team (Euclid Collaboration et al. 2020b) systematically evaluated 9 different machine learning methods on simulated galaxy data and found that most can provide reliable redshift estimates. To further improve performance, Z. Luo et al. (2024b) proposed an LSTM-based model for the CSST survey, which captures sequential correlations in multi-band data and uses Monte Carlo dropout to generate probability distribution functions (PDFs), thereby enhancing reliability. Additionally, studies by X. Zhou et al. (2022) and B. Henghes et al. (2022) demonstrated that multimodal approaches combining photometric and image data (e.g., hybrid MLP-CNN models) can further refine redshift and PDF estimation accuracy.

However, research on quasar photometric redshift estimation remains relatively limited. Compared to galaxies, quasar photometric redshifts generally exhibit lower accuracy and reliability, primarily due to their distinctive physical characteristics: quasar SEDs arise from a combination of power-law continuum emission, broad-line region clouds, dusty tori, and host galaxy light, leading to complex spectral shapes. Moreover, quasars lack pronounced features such as the Balmer break, the 4000Å break, or the Lyman break that characterize galaxy SEDs. These factors result in a highly nonlinear and non-monotonic relationship between quasar colors and redshifts—a phenomenon known as “redshift–color degeneracy” (G. T. Richards et al. 2001; J. A. Newman & D. Gruen 2022). Consequently, methods that perform well for galaxy redshifts often show significantly reduced effectiveness when applied directly to quasars, facing limitations in both accuracy and robustness.

To improve quasar redshift estimation, researchers have explored diverse machine learning strategies. For example, C. Li et al. (2022) developed models using XGBoost, CatBoost, and Random Forest, finding that a two-step strategy—separating quasars into high- and low-redshift subsets before prediction—outperformed single-model approaches. Their subsequent studies (C. Li et al. 2023, 2024) further indicated that these methods generally achieve higher accuracy for galaxies than for quasars. S. J. Curran et al. (2021) evaluated deep learning (DL), k-nearest neighbors

(kNN), and decision tree regression (DTR) for quasar redshift estimation, concluding that kNN and DL delivered superior performance. [J. Pasquet-Itam & J. Pasquet \(2018\)](#) converted quasar light curves into images and applied a convolutional neural network (CNN) to estimate redshifts for SDSS Stripe 82 quasars. Meanwhile, [L. Yao et al. \(2023\)](#) proposed Q-PreNet, a network that integrates image and photometric data for quasar redshift estimation, showing promising results in single-value redshift prediction.

In the photometric redshift estimation, besides single-value (point) estimates, probability distribution functions (PDFs) also serve as an important form of output. Research shows that relying solely on point estimates and their error ranges often fails to fully capture the uncertainties in redshift estimation, while the complete probability distribution can provide a more reliable foundation for subsequent scientific analysis. For instance, in cosmological parameter estimation, using the full probability distribution of redshifts instead of a single value has been proven to effectively improve measurement accuracy ([R. Mandelbaum et al. 2008](#); [A. D. Myers et al. 2009](#)). Therefore, PDFs are particularly important in precision cosmological studies such as weak gravitational lensing and galaxy clustering ([C. Bonnett et al. 2016](#)).

Nevertheless, studies on PDF estimation for quasar photometric redshifts are still relatively scarce. Most current machine learning methods provide only single-point estimates and struggle to produce reliable PDFs, limiting their utility in precision cosmology. This gap is now being addressed by the mixture density network (MDN; [C. M. Bishop 1994](#)). By combining the nonlinear fitting capability of neural networks with the probabilistic framework of Gaussian mixture models (GMM), MDN can directly learn and output complex conditional probability distributions, providing both accurate point estimates and associated uncertainties. For instance, [A. D'Isanto & K. L. Polsterer \(2018\)](#) integrated MDN with a deep convolutional network to propose the DCMDN architecture, which effectively estimates redshift PDFs from quasar images. [C. Zhang et al. \(2024\)](#) further enhanced this approach through cross-modal fusion of photometric attributes (e.g., magnitudes and colors) with image features, achieving excellent PDF estimation performance using a CNN-based residual network combined with MDN.

Although integrating image data through CNNs and MDNs can enhance the performance of photometric redshift estimation, this approach demands substantially more computational resources than methods utilizing only photometric attributes (e.g., flux, magnitude, and color). Data preparation also presents challenges, as the collection and preprocessing of images require significant time and storage space. Furthermore, astronomical images often contain background noise and contaminants, which can adversely affect model performance. Consequently, image-based techniques are generally only feasible for smaller datasets and are difficult to scale to the vast data volumes anticipated from modern sky surveys.

In this study, we propose a novel end-to-end deep learning model, LSTM-MDNz, designed to deliver high-precision point estimates and PDFs for quasar photometric redshifts. The model combines long short-term memory (LSTM) networks with a mixture density network (MDN) and uses multi-band photometric fluxes (and their errors) as input, eliminating the need for manual feature engineering such as color indices. By arranging photometric fluxes and errors in wavelength order as sequential inputs, the LSTM module automatically captures inter-band dependencies and spectral feature correlations. This approach avoids the subjectivity and redundancy of manual feature selection, significantly streamlining input processing.

This study systematically trained and validated the LSTM-MDNz model using a sample of over 550,000 spectroscopically confirmed quasars. By integrating photometric data from four major surveys—the Sloan Digital Sky Survey (SDSS), the DESI Legacy Imaging Surveys (DESI-LS), the Wide-field Infrared Survey Explorer (WISE), and the Galaxy Evolution Explorer (GALEX)—we constructed a comprehensive multi-band dataset covering 14 bands from ultraviolet to infrared. Innovatively, photometric errors were incorporated as input features during modeling to enhance estimation performance. To further evaluate the contribution of different bands to redshift estimation, we conducted in-depth band-ablation experiments to quantitatively analyze the impact of various band combinations on measurement accuracy. The experimental results demonstrate the critical role of ultraviolet and infrared data in mitigating color degeneracy and reducing systematic biases. Furthermore, while maintaining high redshift estimation accuracy, the LSTM-MDNz model exhibited exceptional computational efficiency, characterized by low resource consumption and fast training speed, indicating its strong applicability and scalability for large-scale astronomical data processing.

The remainder of this paper is organized as follows: Section 2 describes the multi-band photometric data and sample construction; Section 3 details the network architecture and training strategy of the LSTM-MDNz model; Section 4 presents experimental results and analysis, including model performance evaluation, band-ablation studies, and key findings; and Section 5 summarizes the research outcomes and provides discussion.

2. DATA

The dataset utilized in this study is derived from the 16th Data Release Quasar Catalog of the Sloan Digital Sky Survey (SDSS DR16Q; B. W. Lyke et al. 2020). This catalog encompasses 750,414 spectroscopically confirmed quasars, each accompanied by celestial coordinates (Right Ascension and Declination), spectroscopic identifiers (MJD, PLATE, FIBERID), spectroscopic redshifts (`spec-z`), and multi-band photometric data from cross-match with other survey programs.

For each quasar in the SDSS DR16Q catalog, we retrieved its photometric data across the five optical bands (u , g , r , i , z) from the SDSS DR16 photometric catalog through coordinate-based cross-matching. To enhance optical wavelength coverage, this study also incorporated photometric measurements in the g , r , and z bands from the DESI Legacy Imaging Surveys Data Release 9 (DESI-LS DR9; D. Schlegel et al. 2021), based on the official online cross-matching catalog between SDSS DR16Q and DESI-LS DR9²³, which adopts a matching radius of 1.5 arcseconds. While these bands partially overlap with those from SDSS, the differences in filter systems provide complementary color information for our machine learning model. This also allows us to evaluate its generalization capability when SDSS u - and i -band data are not available.

The official cross-match catalog between SDSS DR16Q and DESI-LS DR9 is divided into the Northern and Southern galactic caps, with some overlap between the two regions: the Northern cap includes 361,449 quasars, and the Southern cap 430,727. Given the consistency in observational strategy, depth, and data quality between the two caps in DESI-LS, we merged the two datasets to simplify processing. Objects appearing in both caps were retained as independent observational entries to maximize data usage. Tests confirmed that key performance metrics derived from the merged sample were consistent with those from the individual cap samples, indicating that the merging process did not significantly affect model performance.

DESI DR9 also provides forced-photometry infrared fluxes in the $W1$ – $W4$ bands, measured on coadded unWISE images (E. F. Schlafly et al. 2019) using the TRACTOR package (D. Lang et al. 2016). This forced-photometry approach recovers infrared fluxes for faint sources by leveraging the precise optical positions from DESI, thereby substantially improving sample completeness compared to standard WISE catalog-based methods (e.g., ALLWISE or CatWISE).

The final training sample includes model fluxes and associated errors in 14 bands: the five SDSS optical bands (u , g , r , i , z); the three DESI-LS optical bands (g , r , z); the two ultraviolet bands (FUV , NUV) from GALEX, as included in SDSS DR16Q; and the four WISE infrared bands. This multi-wavelength dataset, covering ultraviolet to infrared, supports a comprehensive analysis of quasar spectral energy distributions.

Due to the small number of very high-redshift ($z > 5$) quasars—insufficient for robust statistical training—we restricted the sample to $z \leq 5$, following the approach of A. D’Isanto & K. L. Polsterer (2018) and L. Yao et al. (2023). We also excluded sources with unreliable spectroscopic redshifts (i.e., $z_{\text{warning}} \neq 0$). After filtering, the final sample consisted of 558,111 quasars with complete photometry across all 14 bands. Figure 1 shows the redshift distribution of the sample.

3. METHODOLOGY

In this section, we introduce LSTM-MDNz, a novel end-to-end deep learning model specifically designed for estimating quasar photometric redshifts. The model integrates LSTM networks with MDNs, requiring only multi-band photometric fluxes and their corresponding errors as input. It simultaneously provides high-precision photometric redshift point estimates and complete PDFs. We begin by separately introducing the basic architecture of LSTM recurrent neural networks and MDNs, followed by a detailed description of the overall network design and training methodology for LSTM-MDNz.

3.1. Long Short-Term Memory (LSTM) Network

The LSTM network is a specialized type of recurrent neural network (RNN) first introduced by S. Hochreiter & J. Schmidhuber (1997). It has since been extensively extended and optimized, becoming a mainstream model for processing sequential data. By incorporating a memory cell state, the LSTM effectively mitigates the vanishing and exploding gradient problems that traditional RNNs often encounter with long sequences. Owing to its strong memory

² <https://portal.nersc.gov/cfs/cosmo/data/legacysurvey/dr9/north/external/>

³ <https://portal.nersc.gov/cfs/cosmo/data/legacysurvey/dr9/south/external/>

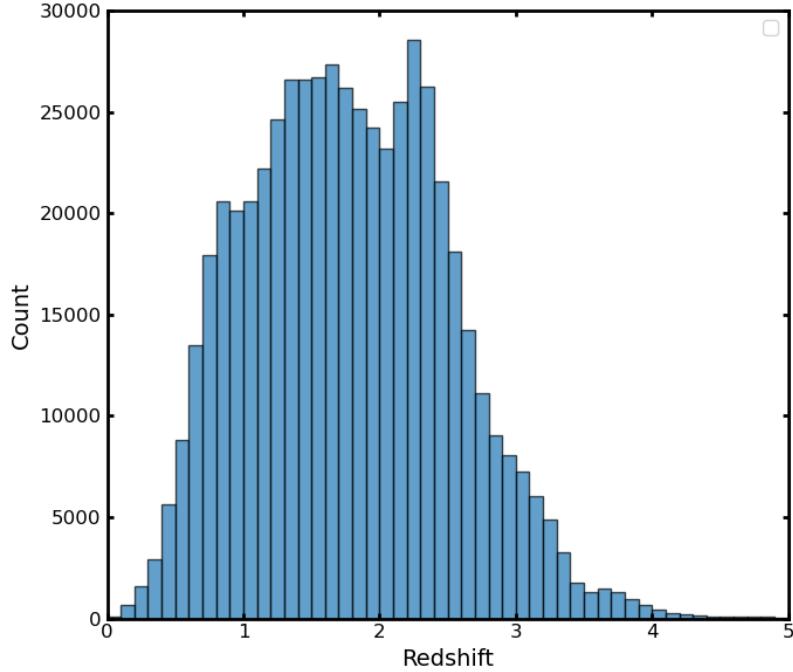


Figure 1. Redshift distribution of the quasar sample with complete 14-band photometry. The vertical axis shows the number of sources, and the horizontal axis the redshift.

capacity and ability to capture long-range dependencies, LSTM has been widely applied in areas such as time series forecasting, reliability prediction, and natural language processing (I. Sutskever et al. 2014; L. Hu et al. 2022; S. S. Tabasi et al. 2023). In recent years, LSTM and other RNN architectures have also gained traction in astronomical research, demonstrating particular promise in photometric redshift estimation (Z. Luo et al. 2024b; G. Teixeira et al. 2024).

The fundamental unit of an LSTM shares structural similarities with a traditional RNN but employs a more sophisticated gating mechanism. Each LSTM unit contains a memory block, which maintains and updates information across time steps via three gates: the forget gate, the input gate, and the output gate, along with a cell state. These components work together to regulate information flow in a fine-grained manner.

Specifically, at time step t , the LSTM unit takes the current input x_t and the previous hidden state h_{t-1} , and computes the following:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad (2)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad (3)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C), \quad (4)$$

where f_t , i_t and o_t represent the forget, input, and output gates, respectively; \tilde{C}_t is the candidate cell state; σ is the sigmoid function; and \tanh is the hyperbolic tangent function. The weight matrices W_f , W_i , W_o , W_C and bias terms b_f , b_i , b_o , b_C are shared across time steps.

The cell state C_t and hidden state h_t are then updated as:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad (5)$$

$$h_t = o_t \odot \tanh(C_t), \quad (6)$$

where \odot denotes element-wise multiplication. Initial states are typically set as $C_0 = 0$ and $h_0 = 0$.

In this mechanism, the forget gate controls how much of the previous cell state is retained; the input gate regulates the integration of new candidate information; and the output gate governs the exposure of the cell state to the hidden output. This gating system allows the LSTM to maintain long-term dependencies and process long sequences effectively, mitigating gradient issues common in vanilla (basic) RNNs.

Compared to conventional machine learning methods, LSTM can automatically learn complex temporal dependencies without manual feature engineering. They excel at capturing long-range patterns and dynamic evolution in sequential data. For example, in natural language processing, LSTMs effectively model long-distance semantic relationships, which improves performance in tasks such as machine translation (S. Hochreiter & J. Schmidhuber 1997; A. Graves et al. 2013). Similarly, in photometric redshift estimation, using LSTM to model multi-band photometric sequences allows the model to capture nonlinear relationships across bands and spectral variations with wavelength, thereby improving prediction accuracy and robustness.

In this study, we employ an LSTM to model multi-band photometric sequences for redshift estimation. This approach enables the model to capture nonlinear relationships between different bands and the variation of flux with wavelength, which is expected to improve the accuracy and stability of predictions. In contrast, a MLP treats the input as an unordered set of features and struggles to utilize the natural order of wavelengths. This limitation hinders its ability to identify local features (e.g., absorption lines) and global trends (e.g., continuum slopes) (Z. Luo et al. 2024b).

To further enhance context capture, we consider a bidirectional LSTM (Bi-LSTM), an extension of the standard LSTM. A Bi-LSTM processes the input sequence in both forward and reverse directions using two separate hidden layers. This architecture allows the model to incorporate both past and future context at every time step, capturing bidirectional dependencies. Bi-LSTMs often improve performance in sequence modeling tasks such as time series forecasting and language understanding (A. Graves & J. Schmidhuber 2005; Z. Huang et al. 2015) and have shown promise in astronomical applications (Z. Luo et al. 2024b). Leveraging these advantages, the LSTM-MDNz model proposed in this paper adopts a Bi-LSTM architecture.

3.2. Mixture Density Network (MDN)

The mixture density network (MDN), introduced by C. M. Bishop (1994), is a neural network architecture designed for regression tasks where the relationship between inputs and outputs may be multimodal or inherently uncertain. Unlike conventional regression models that output a single deterministic value, an MDN models the full conditional probability distribution $p(y|x)$, allowing it to capture complex uncertainty patterns and multiple plausible outcomes for a given input.

An MDN combines the expressive power of a neural network with the flexibility of a mixture model—typically a Gaussian mixture model (GMM). Instead of directly predicting the target variable, the network outputs the parameters of a mixture distribution. For a univariate output y and K Gaussian components, the MDN predicts three sets of parameters per input x : the mixture weights w_k , the means μ_k , and the variances σ_k^2 of each Gaussian component. The mixture weights must be non-negative and sum to 1, the means determine the central position of each component, and the variances control the dispersion of the distributions.

Thus, the output layer must produce $3K$ values. The resulting conditional density is given by:

$$p(y | x) = \sum_{k=1}^K w_k(x) \mathcal{N}(y | \mu_k(x), \sigma_k^2(x)), \quad (7)$$

where \mathcal{N} denotes the Gaussian probability density function. All parameters - the weights w_k , means μ_k , and variances σ_k^2 - are functions of the input x , learned by the neural network.

Architecturally, an MDN generally consists of a feature-extracting backbone (e.g., a multilayer perceptron, MLP) followed by a final linear layer that outputs the $3K$ distribution parameters. To enforce constraints: the mixture weights w_k are passed through a softmax activation to ensure they are positive and sum to 1, the means μ_k are typically unbounded and emitted directly, and the variances σ_k^2 must be positive; this is usually achieved by applying an exponential activation to the corresponding network outputs.

Training is performed via maximum likelihood estimation. For a ground-truth pair (x, y) , the loss is the negative log-likelihood:

$$\mathcal{L} = -\log p(y | x). \quad (8)$$

Minimizing this loss over the training set encourages the MDN to match the predicted distribution to the true data distribution.

Due to its ability to output complete probability distributions, the MDN excels in tasks that require the quantification of predictive uncertainty. In recent years, it has garnered significant attention in the field of astronomical photometric redshift estimation, with relevant studies demonstrating its potential in handling data uncertainty and providing reliable redshift estimates (A. D’Isanto & K. L. Polsterer 2018; C. Zhang et al. 2024; G. Teixeira et al. 2024).

3.3. LSTM-Augmented Mixture Density Network (LSTM-MDNz)

To effectively estimate quasar redshifts from multi-band photometric data and provide reliable uncertainty quantification, this study draws inspiration from the hybrid architecture proposed by A. D’Isanto & K. L. Polsterer (2018) and develops a deep fusion model that integrates LSTM with MDN, referred to as LSTM-MDNz. In contrast to previous approaches, our model does not utilize quasar image data but instead relies solely on multi-band photometric sequences for redshift inference. This model leverages the strength of LSTMs in handling long-term dependencies and dynamic feature extraction in sequential data while incorporating the capability of MDNs for modeling complex conditional probability distributions. This enables the model to directly output comprehensive PDFs of redshift values based on multi-band photometric data, rather than providing a single point estimate. This design effectively addresses uncertainties and multi-modal issues in redshift estimation, significantly enhancing the interpretability and practical value of the results.

As illustrated in Figure 2, the LSTM-MDNz model consists of two core components: a deep bidirectional LSTM feature extractor and an MDN output layer. The feature extractor includes two stacked bidirectional LSTM layers that process the multi-band photometric sequences in a hierarchical manner. The first Bi-LSTM layer captures local dependencies and basic temporal patterns, while the second layer learns more abstract representations and long-range contextual relationships. This deep bidirectional encoding structure enables the model to integrate both forward and backward information, enhancing its ability to perceive spectral energy distribution (SED) shapes and their variations, thereby comprehensively modeling complex dependencies in the input photometric sequences.

The final hidden states from the last Bi-LSTM layer are aggregated into a unified feature vector, which is then passed through a fully connected layer for nonlinear transformation and further abstraction. This high-level feature vector serves as input to the MDN layer. Based on a Gaussian mixture model, the MDN outputs three sets of parameters for each component k : the mixture weight w_k , mean μ_k , and standard deviation σ_k . These parameters are used to reconstruct the conditional probability distribution $p(z|x)$ of the redshift z , enabling probabilistic inference and uncertainty-aware redshift estimation.

3.4. Training the LSTM-MDNz

During model training, the photometric data for each quasar are first transformed into a vector sequence. Specifically, for each band, the measured flux $Flux_t$ and its corresponding error $Flux_t^{err}$ are combined into a two-dimensional vector $(Flux_t, Flux_t^{err})$. These vectors are then arranged in ascending order of wavelength to form the input sequence for the LSTM-MDNz model.

To improve training stability, each feature dimension (i.e., flux or error values for a particular band across all training samples) is standardized as follows:

$$x^* = \frac{x - \mu}{\sigma}, \quad (9)$$

where x is the original value, and μ and σ are the mean and standard deviation of that feature computed over the training set. This standardization ensures that all input features are on a comparable scale, which aids gradient-based optimization and model convergence.

The standardized sequence is then passed to the LSTM encoding module (see Figure 2), which consists of two bidirectional LSTM layers. The first Bi-LSTM layer contains 128 hidden units per direction, and the second contains 64 units per direction. Each Bi-LSTM layer is followed by a dropout layer with a rate of 0.3 and a batch normalization layer. Dropout serves as a regularization technique to reduce overfitting by randomly disabling neuron connections during training (G. E. Hinton et al. 2012; W. Zaremba et al. 2014). Batch normalization stabilizes and accelerates

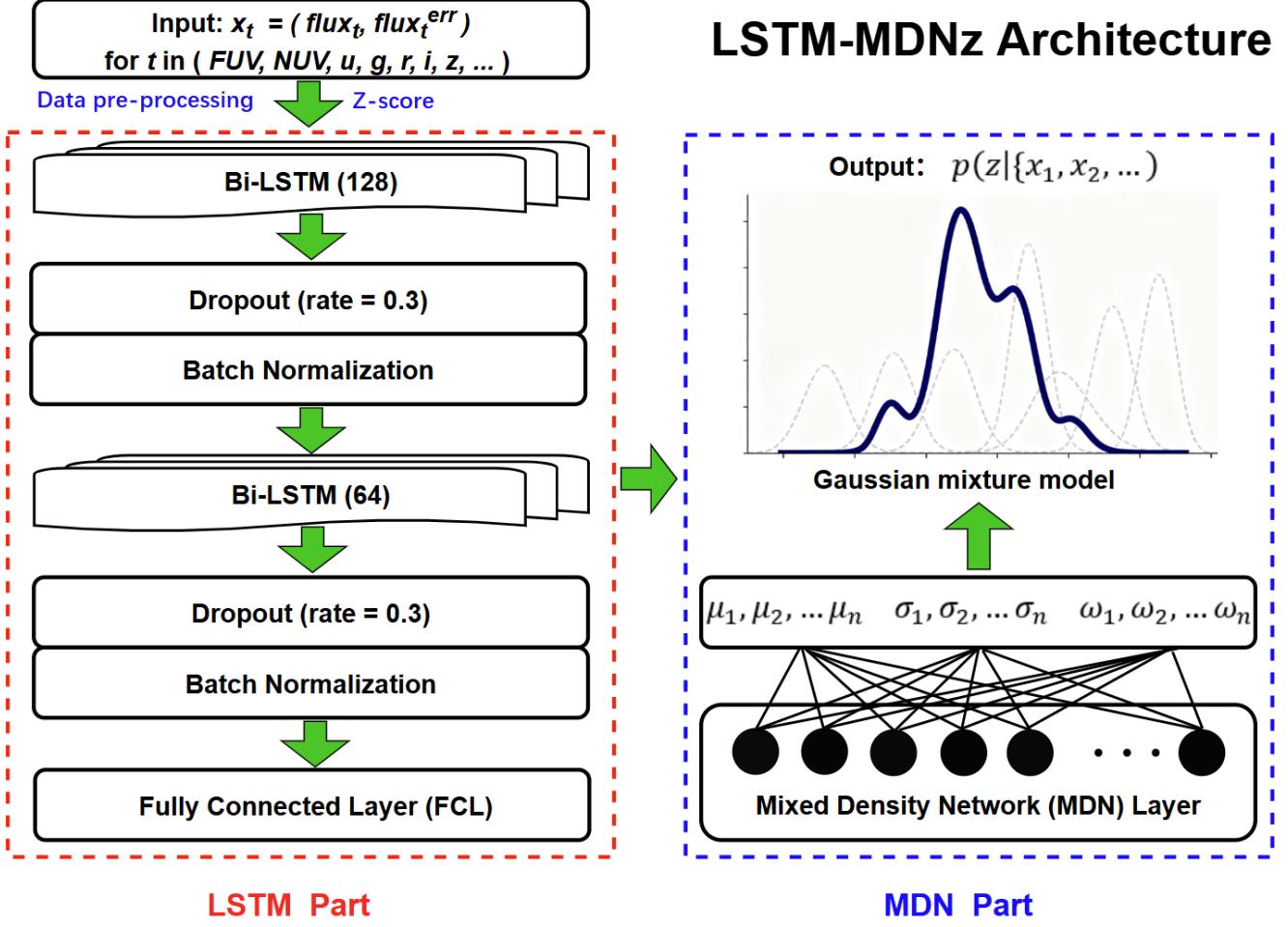


Figure 2. Architecture and data flow of the LSTM-MDNz model. The input consists of multi-band photometric fluxes and their corresponding errors. These sequences are processed by the LSTM encoding module (red dashed box), which includes bidirectional LSTM layers, dropout, and batch normalization for feature extraction and stabilization. The resulting features are transformed via a fully connected layer before being passed to the MDN module (blue dashed box), where a Gaussian mixture model is used to compute the conditional probability density function (PDF) of the photometric redshift.

training by normalizing layer inputs within each mini-batch (S. Ioffe & C. Szegedy 2015). The output of the final batch normalization layer is aggregated into a feature vector and passed through a fully connected layer before entering the MDN module.

The MDN module models the conditional probability distribution $P(z|x)$ of the redshift z given the photometric sequence $x = x_1, x_2, \dots, x_T$ using a Gaussian mixture model. The number of Gaussian components is set to 15, which was determined through systematic validation experiments comparing different counts (ranging from 5 to 25). This value optimally balances model flexibility and generalization performance, yielding the lowest negative log-likelihood on the validation set without overfitting. The network is trained by minimizing the negative log-likelihood loss (Equation 8), which jointly optimizes the LSTM feature extractor and the MDN parameters to accurately fit the redshift distribution.

The model is implemented using the TensorFlow and Keras frameworks. Training is performed on an NVIDIA L40S GPU with a batch size of 128. We use the Adam optimizer (D. P. Kingma 2014) with hyperparameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and an initial learning rate of 0.001. The learning rate is reduced by a factor of 10 every 50 epochs. Early stopping is applied if the validation loss does not improve for 20 consecutive epochs, with the best model weights being restored. The full training process using all bands takes approximately 75 minutes to complete. To ensure

full reproducibility and facilitate community use, the source code, training scripts, and associated data are publicly available in our GitHub repository⁴.

4. EXPERIMENTS AND RESULTS

This section details the training and evaluation of the LSTM-MDNz model using the quasar sample described in Section 2, with the aim of assessing its accuracy and reliability in photometric redshift estimation. The full sample is partitioned into training, validation, and test sets using a 7:1.5:1.5 ratio. Model training uses only the training set, while the validation set is used for real-time performance monitoring during training. Early stopping is triggered if the validation loss shows no improvement over 20 consecutive epochs to prevent overfitting. All final evaluations are performed exclusively on the test set.

4.1. Evaluation Metrics

The primary goal of our model is to produce high-precision point estimates of quasar photometric redshifts along with accurate PDFs. Accordingly, the evaluation metrics are divided into two categories: point estimate statistics, which assess the accuracy of the predicted spectroscopic redshift values, and PDF evaluation metrics, which gauge the reliability and statistical quality of the predicted probability density functions.

For point estimate statistics, we employ several widely adopted metrics: mean squared error (MSE), mean absolute error (MAE), catastrophic outlier fraction (f_{out}), normalized median absolute deviation (σ_{NMAD}), and bias (bias). MSE and MAE are standard regression metrics that quantify the average discrepancy between the true and estimated values:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (z_{\text{spec}}^i - z_{\text{phot}}^i)^2, \quad (10)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |z_{\text{spec}}^i - z_{\text{phot}}^i|, \quad (11)$$

where N is the total number of samples, z_{spec}^i is the spectroscopic redshift (true value) of the i -th sample, and z_{phot}^i is the corresponding photometric redshift estimate.

The catastrophic outlier fraction f_{out} measures the proportion of severely incorrect estimates. Following S. Fotopoulos & S. Paltani (2018) and Euclid Collaboration et al. (2020b), an estimate is classified as a catastrophic outlier if:

$$\frac{|z_{\text{spec}} - z_{\text{phot}}|}{1 + z_{\text{spec}}} > 0.15. \quad (12)$$

Such outliers indicate that the corresponding photometric redshift estimate deviates significantly from the true value.

The normalized median absolute deviation (σ_{NMAD}) evaluates the overall accuracy of the estimates and is defined as (G. B. Brammer et al. 2008):

$$\sigma_{\text{NMAD}} = 1.48 \times \text{median} \left(\frac{|\Delta z|}{1 + z_{\text{spec}}} \right), \quad (13)$$

where $\Delta z = z_{\text{phot}} - z_{\text{spec}}$. σ_{NMAD} is preferred over standard deviation due to its robustness to outliers. The factor 1.48 ensures that under a normal distribution, σ_{NMAD} corresponds to the standard deviation.

Bias quantifies the systematic deviation in the estimates, indicating an overall tendency to over- or underestimate the true redshifts:

$$\text{bias} = \text{median} \left(\frac{z_{\text{phot}} - z_{\text{spec}}}{1 + z_{\text{spec}}} \right). \quad (14)$$

This metric helps identify systematic overestimation or underestimation.

For PDF assessment, we use the continuous ranked probability score (CRPS) and the probability integral transform (PIT) to evaluate the consistency between the predicted and true distributions (A. D'Isanto & K. L. Polsterer 2018). CRPS is a proper scoring rule that measures the accuracy of probabilistic forecasts by comparing the predicted

⁴ <https://github.com/zjluo-code/LSTM-MDNz>

cumulative distribution function (CDF) against the observed value. For a predictive CDF F and an observation x , CRPS is defined as:

$$\text{CRPS}(F, x) = \int_{-\infty}^{+\infty} [F(y) - \mathbf{1}_{\{y \geq x\}}]^2 dy, \quad (15)$$

where y represents a variable in the integral, and $\mathbf{1}_{\{y \geq x\}}$ is the Heaviside indicator function.

PIT assesses the calibration quality of probabilistic forecasts by evaluating how well the predicted distribution represents the true uncertainty (A. P. Dawid 1984; A. D’Isanto & K. L. Polsterer 2018; Euclid Collaboration et al. 2020b; S. Mucesh et al. 2021). For each object, the PIT value is the predicted CDF evaluated at the true redshift:

$$\text{PIT} = \int_{-\infty}^{z_{\text{spec}}} \text{PDF}(z) dz, \quad (16)$$

where z_{spec} is the true redshift, and $\text{PDF}(z)$ is the probability density function at redshift z .

For well-calibrated PDFs, the PIT values across the sample should follow a standard uniform distribution $U(0, 1)$ (A. P. Dawid 1984). This arises because if the PDF is correctly calibrated, the true redshift can be viewed as a random draw from it. Deviations from uniformity in the PIT distribution indicate systematic biases: a U-shaped distribution suggests under-dispersed PDFs, while an inverted U-shape implies over-dispersion. Slope variations in the PIT distribution may also reveal systematic errors (K. L. Polsterer et al. 2016). PIT has been widely used in redshift PDF validation studies (e.g., R. Bordoloi et al. 2010; M. Tanaka et al. 2017; S. Mucesh et al. 2021).

4.2. Model Performance with All Available Bands

This subsection presents the performance of the LSTM-MDNz model when all 14 photometric bands are available. This configuration serves as the benchmark scenario in our study, aimed at evaluating the model’s optimal capability given complete multi-band photometric information. The results will provide a baseline for subsequent band-ablation experiments.

The input features—the photometric flux and its corresponding error for each band—are combined into two-dimensional vectors, which are then arranged in ascending order of wavelength to form a sequence of length 14. In the case of overlapping coverage between the SDSS and DESI-LS bands, the SDSS data is placed before the DESI-LS data in the sequence. Experiments indicate that the specific ordering of bands has negligible effects on the model output. The model architecture follows the description in Section 3.4, utilizing a two-layer bidirectional LSTM for feature extraction and employing a MDN with 15 Gaussian components to output the conditional redshift distribution. The final photometric redshift point estimate (z_{phot}) is taken as the mean of the predicted PDF.

Table 1 summarizes the point estimation performance on the training, validation, and test sets. Under full-band conditions, the model achieves excellent accuracy on the test set: the normalized median absolute deviation (σ_{NMAD}) is 0.037, the catastrophic outlier fraction (f_{out}) is 3.4%, and the bias is nearly zero (-0.0006). The consistency of metrics across all datasets indicates strong generalization without overfitting.

Table 1. Performance metrics of the LSTM-MDNz model for photometric redshift point estimation using all 14 photometric bands.

Data set	No. of objects	σ_{NMAD}	$f_{\text{out}}(\%)$	bias	MSE	MAE
Training Set	390,677	0.036	3.2	-0.0004	0.034	0.105
Validation set	83,717	0.038	3.5	-0.0006	0.040	0.112
Test set	83,717	0.037	3.4	-0.0006	0.039	0.110

Figure 3 illustrates a scatter density distribution, comparing photometric redshifts with spectroscopic redshifts for samples in the test set. This comparison provides a visual assessment of the accuracy and reliability of photometric redshift estimates. As shown in the figure, most data points align closely along the diagonal, indicating strong agreement between predicted (photometric) and true (spectroscopic) redshift values. Importantly, this high degree of consistency is maintained across the full redshift range ($0 \leq z \leq 5$), with no discernible systematic deviations. This consistent performance demonstrates the robustness and accuracy of our photometric redshift estimation method.

The distribution of the CRPS across the entire test set is shown in Figure 4, with a mean value of approximately 0.0783. CRPS is calculated by integrating the squared difference between the CDF of the predicted probabilities and

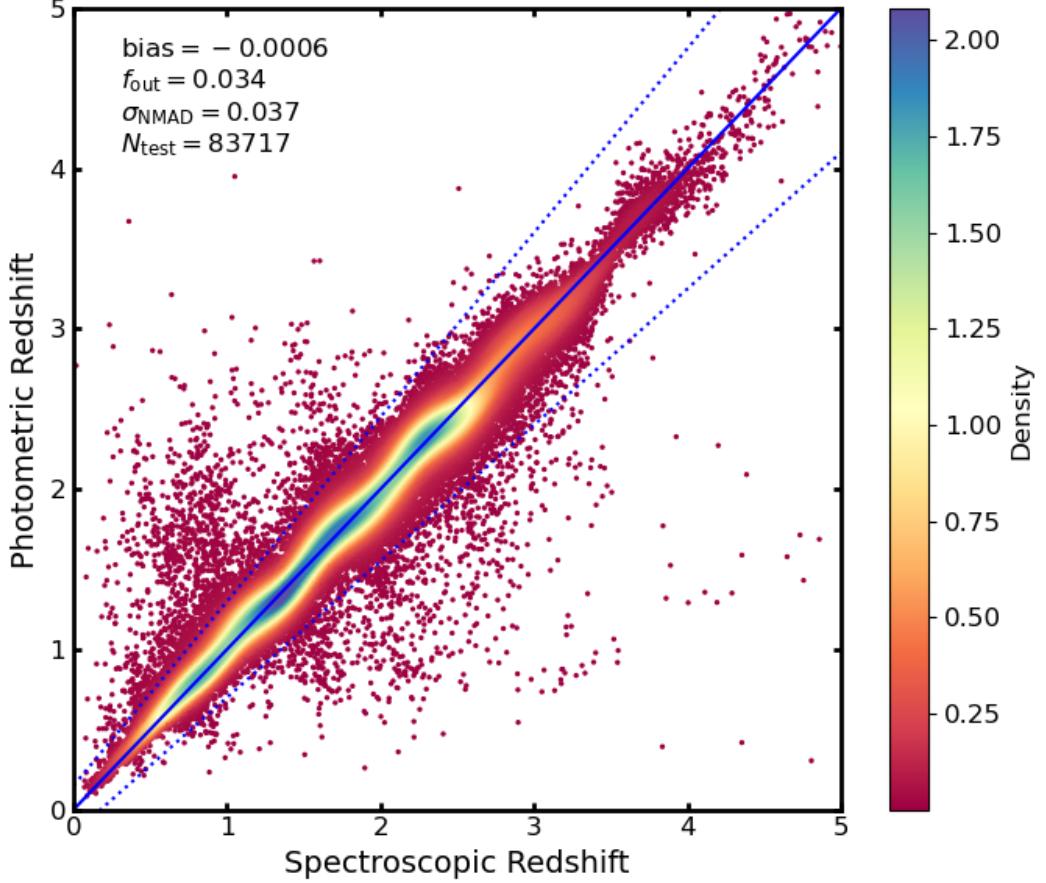


Figure 3. Photometric vs. spectroscopic redshifts for test set samples under full-band (14 bands) input. The solid line is the diagonal; dashed lines show $|z_{\text{spec}} - z_{\text{phot}}|/(1 + z_{\text{spec}}) = 0.15$. Color indicates point density.

the Heaviside indicator function of the true redshift value, as defined in Equation 15. This relatively low mean value indicates that the predicted PDFs closely align with the observed spectroscopic redshift values, thereby validating the model’s capability in accurately assessing the sharpness of the PDFs.

As depicted in Figure 5, we present PDFs (upper panel) and their corresponding CDFs (lower panel) for four randomly selected quasars, drawn from our test set. Each subplot is marked with a red dashed vertical line indicating the true spectroscopic redshift (z_{spec}) of the respective quasar. A detailed examination of the plots reveals that most observed spectroscopic redshift values lie within the high-probability regions of their predicted PDFs. This outcome signifies not only the model’s proficiency in generating precise point predictions but also its exceptional ability to produce reliable uncertainty quantifications. The consistent proximity between predicted and true redshifts corroborates the robustness and accuracy of our methodology.

The left panel of Figure 6 illustrates the PIT distribution for our test set samples. A perfectly calibrated model would produce PIT values uniformly distributed over the interval $[0, 1]$, as indicated by the red horizontal dashed line in the figure. The actual PIT distribution of our samples, shown here as a blue histogram, closely resembles this uniform distribution, with only minor localized deviations.

The right panel of Figure 6 presents a quantile-quantile (Q-Q) plot comparing the observed PIT values to those expected under a uniform $(0, 1)$ distribution. In an ideally calibrated model, all points would lie along the red reference line representing the theoretical quantiles. As demonstrated in this figure, our model exhibits excellent calibration, with most points cluster tightly around the reference line, indicating that the statistical properties of the PIT values are highly consistent with a uniform distribution.

This strong agreement between the observed and expected distributions validates not only the accuracy of our point estimates but also the reliability of the uncertainty quantification provided by the model. The close adherence to

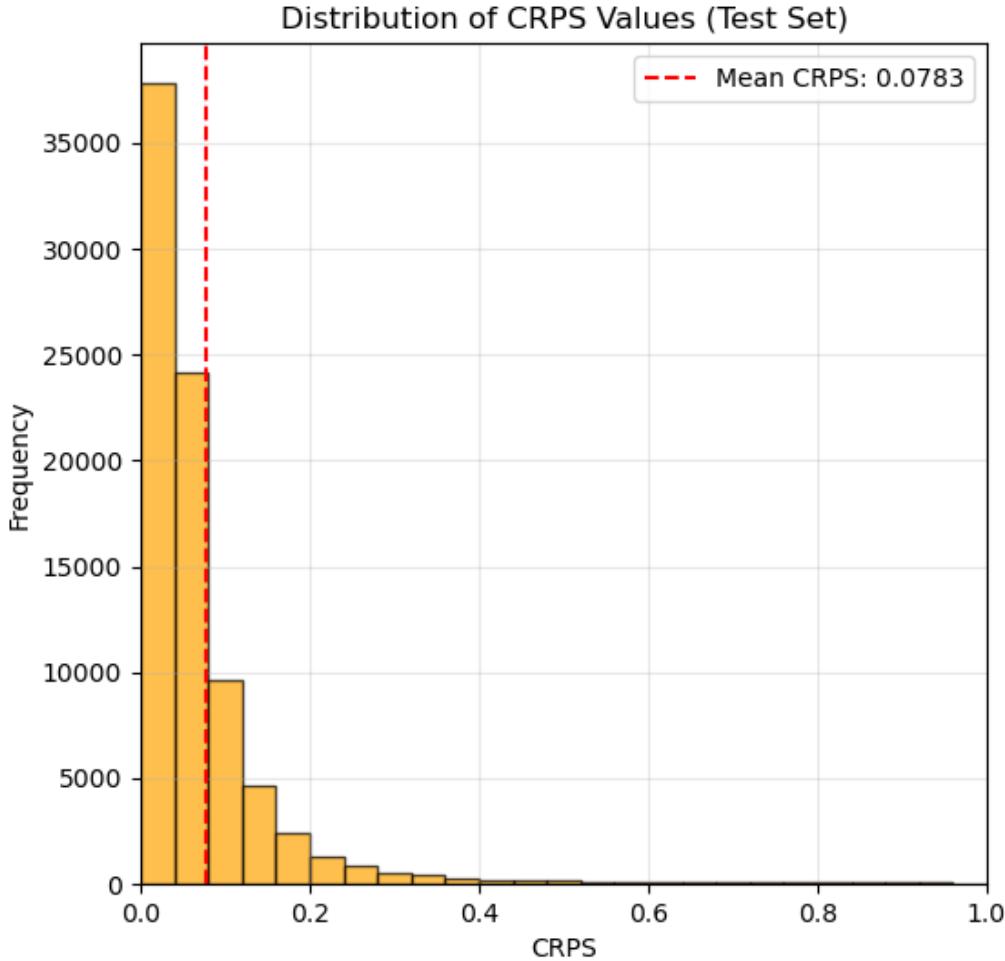


Figure 4. Distribution of CRPS values for the test set under full-band (14 bands) input. The red dashed line marks the mean CRPS.

uniformity in both the histogram and Q-Q plot underscores the robustness of our approach in capturing the inherent uncertainties in redshift estimation.

Our evaluation demonstrates that fusing 14 photometric bands, ranging from ultraviolet to infrared, effectively samples the spectral energy distribution (SED). This enables the model to better capture the characteristics of the SED shape of quasars and to break the redshift-color degeneracy. These advances yield not only more accurate point estimates but also better-calibrated probabilistic predictions. The resulting uncertainties faithfully represent the true error distribution, which is crucial for scientific applications like cosmological parameter inference.

4.3. Band-ablation Experiments

To systematically evaluate the contributions of various band data to the accuracy of quasar photometric redshift estimation and to confirm the critical role of ultraviolet (GALEX) and infrared (WISE) bands in resolving the redshift-color degeneracy, we conducted a series of systematic band-ablation experiments. These experiments were designed to assess how the absence of specific bands affects model performance by examining different combinations of input data while keeping model architecture, hyperparameters, and training strategies consistent.

The experiments employed six distinct band configurations:

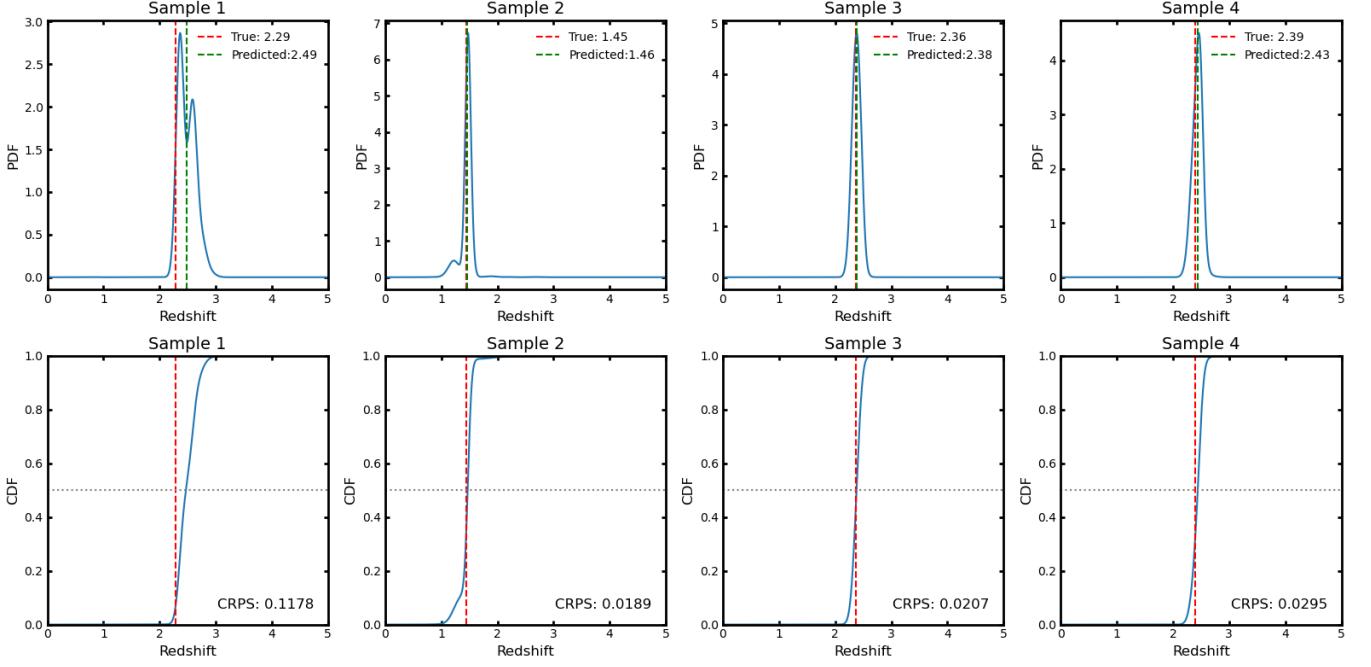


Figure 5. Examples of photometric redshift predictions for four randomly selected quasars under full-band (14 bands) input. The solid blue line in the upper panel represents the predicted PDF, while the solid blue line in the lower panel shows the corresponding CDF. The red dashed line indicates the position of the true spectroscopic redshift (z_{spec}), and the blue dashed line represents the photometric redshift (z_{phot}).

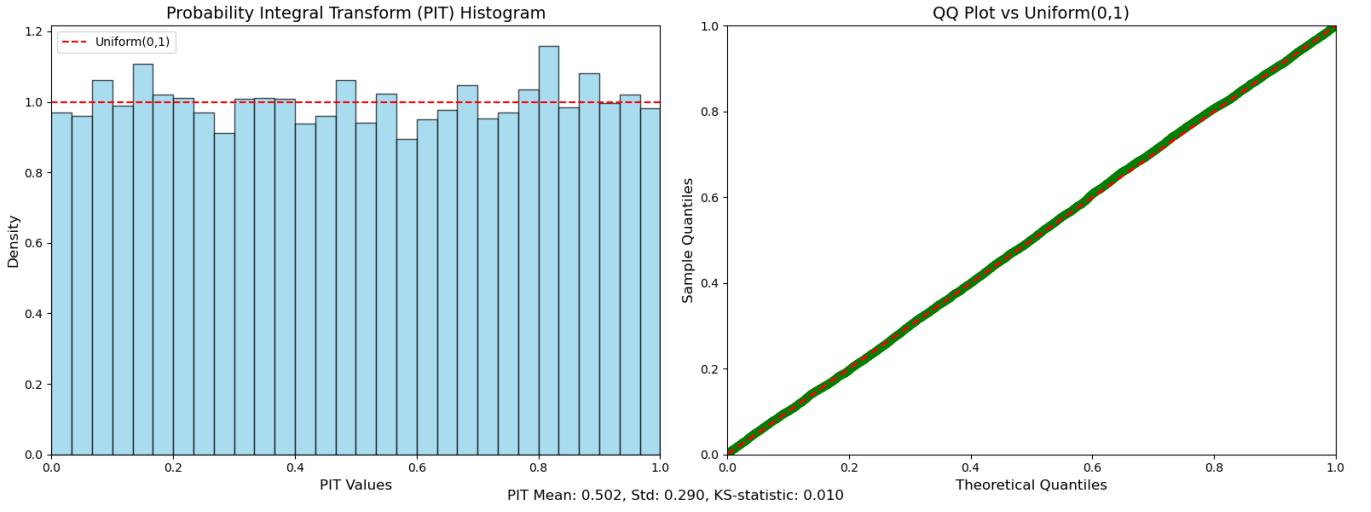


Figure 6. PIT histogram (left) and Q–Q plot (right) for the test set under full-band (14 bands) input. In the left panel, the blue histogram represents the observed PIT distribution, and the red dashed line indicates the reference line for the ideal uniform distribution ($U(0,1)$). In the right panel, the scatter points show the correspondence between the observed quantiles and theoretical quantiles, with the red dashed line representing the theoretical $y = x$ line.

1. All Bands (14 bands): Serving as the baseline, this configuration incorporates data from all available surveys: GALEX ultraviolet (FUV , NUV), SDSS optical (u , g , r , i , z), DESI-LS optical (g , r , z), and WISE infrared ($W1$, $W2$, $W3$, $W4$).

2. Optical Only (8 bands): This setup uses only optical bands from SDSS and DESI-LS, simulating scenarios where ultraviolet and infrared data are unavailable.

3. Optical + WISE (12 bands): Combining optical bands with WISE infrared data while excluding GALEX ultraviolet observations.

4. Optical + GALEX (10 bands): Merging optical bands with GALEX ultraviolet data while omitting WISE infrared contributions.

5. SDSS + WISE (9 bands): Reproducing a combination frequently used in previous studies (L. Yao et al. 2023; C. Zhang et al. 2024), comprising SDSS optical bands and WISE infrared data.

6. DESI + WISE (7 bands): Representing scenarios where specific SDSS filters (particularly u and i bands) are unavailable, testing the capability of DESI Legacy Survey optical data combined with WISE infrared observations.

All experiments employed identical training, validation, and test datasets to ensure result comparability. For point estimation accuracy, we focused on changes in two key metrics: normalized median absolute deviation (σ_{NMAD}) and the catastrophic outlier rate (f_{out}). Regarding probabilistic predictions, our analysis concentrated primarily on the mean CRPS to assess the sharpness and reliability of the predicted PDFs.

Table 2 summarizes the performance of the LSTM-MDNz model on the test set under different band combinations. The experimental results clearly demonstrate the importance of various band data for accurate photometric redshift estimation. When utilizing all 14 bands, the model achieves superior performance, with $\sigma_{\text{NMAD}} = 0.037$, $f_{\text{out}} = 3.4\%$, and mean CRPS = 0.0783. By comparison, when only the 8 optical bands from SDSS+DESI are used, model performance significantly deteriorates: σ_{NMAD} increases to 0.068, f_{out} rises to 20.6%, and the mean CRPS worsens to 0.1657. This result aligns with the redshift-color degeneracy challenge discussed in the introduction, highlighting that optical information alone is insufficient for uniquely determining quasar redshifts.

When infrared or ultraviolet data were added to the optical bands (SDSS + DESI), the model's performance showed significant improvement in both scenarios. Specifically, after incorporating WISE infrared data, the normalized median absolute deviation (σ_{NMAD}) decreased to 0.041, representing an approximate 40% improvement compared to the configuration using only optical bands. Additionally, the catastrophic outlier rate (f_{out}) was dramatically reduced to 5.4%, indicating an approximate 74% improvement, while the CRPS decreased to 0.0911, reflecting around a 45% enhancement.

In the case of adding GALEX ultraviolet data, σ_{NMAD} dropped to 0.047, which corresponds to approximately 31% improvement compared to the optical-only configuration. The outlier rate significantly decreased to 9.2%, equating to around a 55% improvement, and the mean CRPS reduced to 0.1123, indicating approximately a 32% enhancement. These findings demonstrate that both infrared and ultraviolet data provide crucial information not captured by optical bands alone, enabling the model to better distinguish between quasars with similar optical colors but differing redshifts due to degeneracy.

Table 2. Performance of the LSTM-MDNz Model on Photometric Redshift Estimation Across Different Band Combinations

Band combination	No. of bands	σ_{NMAD}	$f_{\text{out}}(\%)$	bias	MSE	MAE	CPRS (mean)
All Bands	14	0.037	3.4	-0.0006	0.039	0.110	0.0783
Optical + WISE	12	0.041	5.4	-0.0005	0.056	0.129	0.0911
Optical + GALEX	10	0.047	9.2	-0.0010	0.086	0.163	0.1123
SDSS + WISE	9	0.052	7.9	-0.0010	0.073	0.159	0.1106
DESI + WISE	7	0.080	16.8	-0.0017	0.137	0.235	0.1635
Optical Only	8	0.068	20.6	0.0050	0.169	0.258	0.1657

Furthermore, a comparison of model performance across different combinations (Table 2) indicates that optimal results are achieved only when all bands (ultraviolet, optical, and infrared) are employed together. The performance advantage of the “All Bands” configuration over the “Optical + WISE” or “Optical + GALEX” combinations demonstrates that the complementary information provided by ultraviolet and infrared bands is not entirely redundant; rather, it exhibits a synergistic effect that collectively enhances the model's accuracy.

Table 2 also illustrates an interesting phenomenon: although both SDSS and DESI-LS belong to the optical band category, their differing filter response functions allow for the integration of complementary optical data from different

survey projects, which can significantly enhance model performance. For instance, compared to the SDSS+WISE combination, the Optical + WISE configuration (i.e., SDSS + DESI + WISE) achieves a remarkable 21% reduction in the normalized median absolute deviation (σ_{NMAD}) from 0.052 to 0.041, a 34% decrease in the outlier rate (f_{out}) from 7.9% to 5.4%, and an 18% reduction in the mean CRPS from 0.1106 to 0.0911. Conversely, when the optical band coverage is reduced—such as by replacing SDSS+WISE with DESI + WISE, which results in the loss of u -band and i -band information—the model’s performance in photometric redshift prediction declines noticeably.

For the SDSS + WISE combination commonly adopted in previous studies, the performance achieved by the LSTM-MDNz model in this paper is $\sigma_{\text{NMAD}} = 0.052$, $f_{\text{out}} = 7.9\%$, and mean CRPS = 0.1106. This result significantly surpasses the performance reported by [S. J. Curran et al. \(2021\)](#), who utilized similar SDSS + WISE photometric data ($\sigma_{\text{NMAD}} = 0.103$, $f_{\text{out}} = 24\%$). The improvement can likely be attributed to our probabilistic LSTM-MDN architecture, which directly models the sequential nature of multi-band data and explicitly quantifies redshift uncertainties through mixture density modelling. It also outperforms the quasar redshift estimation results obtained by [L. Yao et al. \(2023\)](#) using their Q-PreNet model that integrates image and photometric features ($\sigma_{\text{NMAD}} = 0.070$, $f_{\text{out}} = 13.7\%$, mean CRPS = 0.1318). Notably, this advancement is achieved despite our model using only photometric sequences, highlighting the effectiveness of our LSTM-based encoding. Furthermore, our model’s performance is comparable to the PDF estimation results achieved by [C. Zhang et al. \(2024\)](#) through cross-modal fusion of photometric and image features using CNNs and MDNs, with a mean CRPS of 0.1187.

However, when compared to the 14-band fusion scheme proposed in this paper, labeled as All Bands, the SDSS + WISE combination exhibits a significant performance gap. Specifically, the σ_{NMAD} value increases by approximately 41%, the catastrophic outlier rate f_{out} rises by about 132%, and the mean CRPS increases by approximately 41%. This comparison underscores the critical importance of further integrating DESI-LS, which provides complementary optical information, and GALEX, offering ultraviolet data, on top of the SDSS and WISE data to achieve current high-precision photometric redshifts.

Figure 7 illustrates the scatter density distributions of photometric versus spectroscopic redshifts under various band combinations. Observations reveal that as certain bands are excluded, particularly ultraviolet or infrared data, vertical stripe-like patterns in the scatter plots become more pronounced. These stripes indicate increased redshift confusion due to color degeneracy. Additionally, both systematic bias and dispersion of photometric redshifts significantly increase when critical bands are absent, highlighting the importance of comprehensive band coverage for accurate redshift estimation.

Through our comprehensive band-ablation experiments, we have demonstrated that infrared (WISE) and ultraviolet (GALEX) bands play pivotal roles in addressing the redshift–color degeneracy of quasars. These bands not only mitigate systematic biases but also significantly enhance the accuracy and reliability of photometric redshift estimations. In addition, the integration of complementary optical information from diverse survey projects, such as the DESI Legacy Surveys (DESI-LS), provides substantial support for improved model performance.

5. SUMMARY AND DISCUSSION

This study addresses the fundamental challenge of redshift-color degeneracy in quasar photometric redshift estimation by proposing LSTM-MDNz, an end-to-end framework that integrates long short-term memory networks with mixture density networks. The model processes multiband photometric fluxes and their measurement errors to simultaneously generate both precise point estimates and complete probability distribution functions, eliminating manual feature engineering while providing comprehensive uncertainty quantification.

The model’s architecture is specifically designed to capture spectral dependencies by structuring photometric data as wavelength-ordered sequences. The LSTM component extracts spectral energy distribution features and inter-band relationships, while the MDN layer maps these features to parameters of a Gaussian mixture distribution. This integrated approach enables robust probabilistic modeling of redshift uncertainties, effectively addressing limitations inherent in traditional estimation methods.

The LSTM-MDNz model demonstrates good performance on a quasar dataset covering 14 bands from ultraviolet to infrared. In terms of point estimation, it achieves a normalized median absolute deviation (σ_{NMAD}) of 0.037, a catastrophic outlier rate (f_{out}) as low as 3.4%, and a bias close to zero. For probabilistic prediction, statistical analysis of the PIT indicates that the predicted photometric redshift PDFs are well-calibrated, while the CRPS score reveals a high degree of correspondence between the predicted PDFs and the spectroscopically measured redshift values, demonstrating accurate and reliable PDF uncertainty estimation. These results validate the strong capability

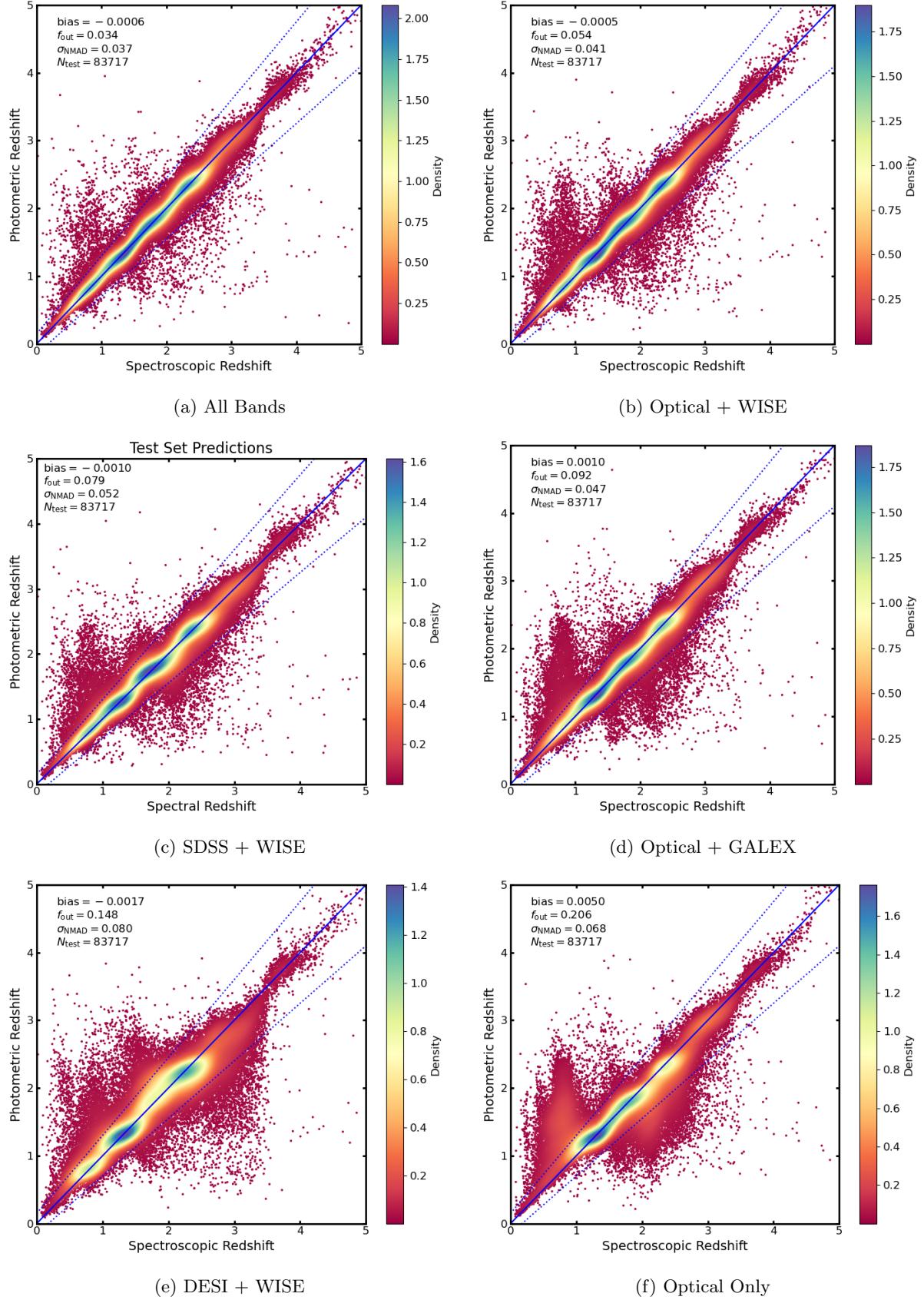


Figure 7. Spectroscopic redshifts versus predicted photometric redshifts derived from the proposed LSTM-MDNz model for the test set under different band combinations.

of the LSTM architecture in processing multi-band photometric sequences and capturing spectral energy distribution features, as well as the effectiveness of the MDN in modeling complex conditional probability distributions.

Through systematic band-ablation experiments, we have quantitatively demonstrated the crucial role of ultraviolet (GALEX) and infrared (WISE) bands in addressing the redshift-color degeneracy issue in quasars. Compared to using optical bands alone, incorporating either ultraviolet or infrared data significantly enhances model performance—reducing σ_{NMAD} by approximately 30-40%, lowering f_{out} (catastrophic outlier rate) by 55-75%, and decreasing CPRS by 30-45%. Additionally, integrating complementary optical band data from different survey projects also yields significant performance gains for the model. The ultimate optimal performance arises from a synergistic effect across all bands (ultraviolet, optical, and infrared) rather than simple additive effects. This underscores the importance of constructing broadly covered multi-band datasets for robust and accurate modeling.

Compared with existing studies, the LSTM-MDNz model in this research has notable advantages. When using similar SDSS and WISE benchmark photometric data, it can match and even surpass the performance of complex models that require multimodal data, which validates the model's good ability to extract key information from photometric data. The core value of the model lies in its end-to-end modeling approach. It relies solely on photometric data and does not require manual feature engineering, thus having great potential for large-scale applications. We believe that LSTM-MDNz provides a powerful and scalable solution to deal with the vast amount of data from next-generation sky surveys such as DESI, LSST, Euclid, and CSST.

While the LSTM-MDNz model demonstrates promising performance in quasar photometric redshift estimation, its capabilities and broader applicability can be further enhanced in several directions. Firstly, the current model is primarily validated on bright SDSS quasars; its performance on fainter sources with larger photometric errors—common in surveys like LSST and Euclid—requires further testing, as increased noise may exacerbate redshift-color degeneracy and affect PDF calibration. Future work will explore noise-aware training or transfer learning on deeper, noisier datasets to improve robustness. Secondly, generalization to higher redshifts ($z > 5$) is currently limited by scarce training data, but targeted optimization can be pursued as high-redshift samples expand. Thirdly, integrating additional data modalities (e.g., variability features or morphological parameters) could further boost accuracy and robustness. Lastly, systematic evaluation under varied observational depths and selection functions will be crucial for optimizing the model's practical utility across diverse surveys.

ACKNOWLEDGMENTS

Z.L. acknowledges the support from the National Natural Science Foundation of China (Grant No. 12573009) and the scientific research grants from the China Manned Space Project with Grand No. CMS-CSST-2025-A07. L.F. acknowledges the support from the Innovation Program of Shanghai Municipal Education Commission (Grant No. 2025GDZKZD04) and China Manned Space Project with Grant No. CMS-CSST-2025-A05. S.Z. acknowledges support from the National Natural Science Foundation of China (Grant No. 12173026), the National Key Research and Development Program of China (Grant No. 2022YFC2807303), the Shanghai Science and Technology Fund (Grant No. 23010503900), the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, and the Shuguang Program (23SG39) of the Shanghai Education Development Foundation and Shanghai Municipal Education Commission. H.X. acknowledges the support from the National Natural Science Foundation of China (NSFC 12203034), the Shanghai Science and Technology Fund (22YF1431500), and the Shanghai Municipal Education Commission regarding artificial intelligence empowered research. This work is also supported by the National Natural Science Foundation of China under Grant No. 12141302.

REFERENCES

- Abbott, T. M. C., Adamów, M., Aguena, M., et al. 2021, ApJS, 255, 20, doi: [10.3847/1538-4365/ac00b3](https://doi.org/10.3847/1538-4365/ac00b3)
- Aihara, H., Arimoto, N., Armstrong, R., et al. 2018, PASJ, 70, S4, doi: [10.1093/pasj/psx066](https://doi.org/10.1093/pasj/psx066)
- Akeson, R., Armus, L., Bachelet, E., et al. 2019, arXiv e-prints, arXiv:1902.05569, doi: [10.48550/arXiv.1902.05569](https://doi.org/10.48550/arXiv.1902.05569)
- Almosallam, I. A., Lindsay, S. N., Jarvis, M. J., & Roberts, S. J. 2016, MNRAS, 455, 2387, doi: [10.1093/mnras/stv2425](https://doi.org/10.1093/mnras/stv2425)

- Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540, doi: [10.1046/j.1365-8711.1999.02978.x](https://doi.org/10.1046/j.1365-8711.1999.02978.x)
- Ball, N. M., Brunner, R. J., Myers, A. D., et al. 2007, ApJ, 663, 774, doi: [10.1086/518362](https://doi.org/10.1086/518362)
- Benítez, N. 2000, ApJ, 536, 571, doi: [10.1086/308947](https://doi.org/10.1086/308947)
- Bishop, C. M. 1994, Aston University.[Google Scholar]. <https://publications.aston.ac.uk/id/eprint/373/>
- Bolzonella, M., Miralles, J. M., & Pelló, R. 2000, A&A, 363, 476, doi: [10.48550/arXiv.astro-ph/0003380](https://doi.org/10.48550/arXiv.astro-ph/0003380)
- Bolzonella, M., Miralles, J.-M., & Pelló, R. 2011,, Astrophysics Source Code Library, record ascl:1108.010 <http://ascl.net/1108.010>
- Bonnett, C., Troxel, M. A., Hartley, W., et al. 2016, PhRvD, 94, 042005, doi: [10.1103/PhysRevD.94.042005](https://doi.org/10.1103/PhysRevD.94.042005)
- Bordoloi, R., Lilly, S. J., & Amara, A. 2010, MNRAS, 406, 881, doi: [10.1111/j.1365-2966.2010.16765.x](https://doi.org/10.1111/j.1365-2966.2010.16765.x)
- Brammer, G. B., van Dokkum, P. G., & Coppi, P. 2008, ApJ, 686, 1503, doi: [10.1086/591786](https://doi.org/10.1086/591786)
- Carrasco Kind, M., & Brunner, R. J. 2013, MNRAS, 432, 1483, doi: [10.1093/mnras/stt574](https://doi.org/10.1093/mnras/stt574)
- Collierier, A. A., & Lahav, O. 2004, PASP, 116, 345, doi: [10.1086/383254](https://doi.org/10.1086/383254)
- Curran, S. J., Moss, J. P., & Perrott, Y. C. 2021, MNRAS, 503, 2639, doi: [10.1093/mnras/stab485](https://doi.org/10.1093/mnras/stab485)
- Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. 2016, MNRAS, 460, 1270, doi: [10.1093/mnras/stw641](https://doi.org/10.1093/mnras/stw641)
- Dawid, A. P. 1984, Journal of the Royal Statistical Society: Series A (General), 147, 278, doi: [10.2307/2981683](https://doi.org/10.2307/2981683)
- de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Experimental Astronomy, 35, 25, doi: [10.1007/s10686-012-9306-1](https://doi.org/10.1007/s10686-012-9306-1)
- D'Isanto, A., & Polsterer, K. L. 2018, A&A, 609, A111, doi: [10.1051/0004-6361/201731326](https://doi.org/10.1051/0004-6361/201731326)
- Euclid Collaboration, Blanchard, A., Camera, S., et al. 2020a, A&A, 642, A191, doi: [10.1051/0004-6361/202038071](https://doi.org/10.1051/0004-6361/202038071)
- Euclid Collaboration, Despres, G., Paltani, S., et al. 2020b, A&A, 644, A31, doi: [10.1051/0004-6361/202039403](https://doi.org/10.1051/0004-6361/202039403)
- Fan, X., Carilli, C. L., & Keating, B. 2006, ARA&A, 44, 415, doi: [10.1146/annurev.astro.44.051905.092514](https://doi.org/10.1146/annurev.astro.44.051905.092514)
- Fernández-Soto, A., Lanzetta, K. M., & Yahil, A. 1999, ApJ, 513, 34, doi: [10.1086/306847](https://doi.org/10.1086/306847)
- Firth, A. E., Lahav, O., & Somerville, R. S. 2003, MNRAS, 339, 1195, doi: [10.1046/j.1365-8711.2003.06271.x](https://doi.org/10.1046/j.1365-8711.2003.06271.x)
- Fotopoulou, S., & Paltani, S. 2018, A&A, 619, A14, doi: [10.1051/0004-6361/201730763](https://doi.org/10.1051/0004-6361/201730763)
- Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996, AJ, 111, 1748, doi: [10.1086/117915](https://doi.org/10.1086/117915)
- Geach, J. E. 2012, MNRAS, 419, 2633, doi: [10.1111/j.1365-2966.2011.19913.x](https://doi.org/10.1111/j.1365-2966.2011.19913.x)
- Gerdes, D. W., Sypniewski, A. J., McKay, T. A., et al. 2010, ApJ, 715, 823, doi: [10.1088/0004-637X/715/2/823](https://doi.org/10.1088/0004-637X/715/2/823)
- Gong, Y., Liu, X., Cao, Y., et al. 2019, ApJ, 883, 203, doi: [10.3847/1538-4357/ab391e](https://doi.org/10.3847/1538-4357/ab391e)
- Graves, A., Mohamed, A.-r., & Hinton, G. 2013, arXiv e-prints, arXiv:1303.5778, doi: [10.48550/arXiv.1303.5778](https://doi.org/10.48550/arXiv.1303.5778)
- Graves, A., & Schmidhuber, J. 2005, Neural Networks, 18, 602, doi: <https://doi.org/10.1016/j.neunet.2005.06.042>
- Green, J., Schechter, P., Baltay, C., et al. 2012, arXiv e-prints, arXiv:1208.4012, doi: [10.48550/arXiv.1208.4012](https://doi.org/10.48550/arXiv.1208.4012)
- Henghes, B., Thiyagalingam, J., Pettitt, C., Hey, T., & Lahav, O. 2022, MNRAS, 512, 1696, doi: [10.1093/mnras/stac480](https://doi.org/10.1093/mnras/stac480)
- Hildebrandt, H., Arnouts, S., Capak, P., et al. 2010, A&A, 523, A31, doi: [10.1051/0004-6361/201014885](https://doi.org/10.1051/0004-6361/201014885)
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. 2012, arXiv e-prints, arXiv:1207.0580, doi: [10.48550/arXiv.1207.0580](https://doi.org/10.48550/arXiv.1207.0580)
- Hochreiter, S., & Schmidhuber, J. 1997, Neural Computation, 9, 1735, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)
- Hopkins, P. F., & Elvis, M. 2010, MNRAS, 401, 7, doi: [10.1111/j.1365-2966.2009.15643.x](https://doi.org/10.1111/j.1365-2966.2009.15643.x)
- Hoyle, B., Rau, M. M., Paech, K., et al. 2015, MNRAS, 452, 4183, doi: [10.1093/mnras/stv1551](https://doi.org/10.1093/mnras/stv1551)
- Hu, L., Chen, X., & Wang, L. 2022, ApJ, 930, 70, doi: [10.3847/1538-4357/ac5c48](https://doi.org/10.3847/1538-4357/ac5c48)
- Huang, Z., Xu, W., & Yu, K. 2015, arXiv e-prints, arXiv:1508.01991, doi: [10.48550/arXiv.1508.01991](https://doi.org/10.48550/arXiv.1508.01991)
- Ioffe, S., & Szegedy, C. 2015, arXiv e-prints, arXiv:1502.03167, doi: [10.48550/arXiv.1502.03167](https://doi.org/10.48550/arXiv.1502.03167)
- Ivezic, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111, doi: [10.3847/1538-4357/ab042c](https://doi.org/10.3847/1538-4357/ab042c)
- Kingma, D. P. 2014, arXiv preprint arXiv:1412.6980
- Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511, doi: [10.1146/annurev-astro-082708-101811](https://doi.org/10.1146/annurev-astro-082708-101811)
- Lang, D., Hogg, D. W., & Mykytyn, D. 2016, Astrophysics Source Code Library, record ascl:1604.008 <http://ascl.net/1604.008>
- Lanzetta, K. M., Yahil, A., & Fernández-Soto, A. 1996, Nature, 381, 759, doi: [10.1038/381759a0](https://doi.org/10.1038/381759a0)
- Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193, doi: [10.48550/arXiv.1110.3193](https://doi.org/10.48550/arXiv.1110.3193)
- Li, C., Zhang, Y., Cui, C., et al. 2022, MNRAS, 509, 2289, doi: [10.1093/mnras/stab3165](https://doi.org/10.1093/mnras/stab3165)
- Li, C., Zhang, Y., Cui, C., et al. 2023, MNRAS, 518, 513, doi: [10.1093/mnras/stac3037](https://doi.org/10.1093/mnras/stac3037)
- Li, C., Zhang, Y., Cui, C., et al. 2024, AJ, 168, 233, doi: [10.3847/1538-3881/ad7c52](https://doi.org/10.3847/1538-3881/ad7c52)

- LSST Science Collaboration: Abell, P. A., Allison, J., Anderson, S. F., et al. 2009, arXiv e-prints, arXiv:0912.0201, doi: [10.48550/arXiv.0912.0201](https://doi.org/10.48550/arXiv.0912.0201)
- Lu, J., Luo, Z., Chen, Z., et al. 2024, MNRAS, 527, 12140, doi: [10.1093/mnras/stad3976](https://doi.org/10.1093/mnras/stad3976)
- Luo, Z., Tang, Z., Chen, Z., et al. 2024a, MNRAS, 531, 3539, doi: [10.1093/mnras/stae1397](https://doi.org/10.1093/mnras/stae1397)
- Luo, Z., Li, Y., Lu, J., et al. 2024b, MNRAS, 535, 1844, doi: [10.1093/mnras/stae2446](https://doi.org/10.1093/mnras/stae2446)
- Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, ApJS, 250, 8, doi: [10.3847/1538-4365/aba623](https://doi.org/10.3847/1538-4365/aba623)
- Mandelbaum, R., Seljak, U., & Hirata, C. M. 2008, JCAP, 2008, 006, doi: [10.1088/1475-7516/2008/08/006](https://doi.org/10.1088/1475-7516/2008/08/006)
- Matsuoka, Y., Onoue, M., Kashikawa, N., et al. 2016, ApJ, 828, 26, doi: [10.3847/0004-637X/828/1/26](https://doi.org/10.3847/0004-637X/828/1/26)
- Mucesh, S., Hartley, W. G., Palmese, A., et al. 2021, MNRAS, 502, 2770, doi: [10.1093/mnras/stab164](https://doi.org/10.1093/mnras/stab164)
- Myers, A. D., White, M., & Ball, N. M. 2009, MNRAS, 399, 2279, doi: [10.1111/j.1365-2966.2009.15432.x](https://doi.org/10.1111/j.1365-2966.2009.15432.x)
- Newman, J. A., & Gruen, D. 2022, ARA&A, 60, 363, doi: [10.1146/annurev-astro-032122-014611](https://doi.org/10.1146/annurev-astro-032122-014611)
- Pasquet, J., Bertin, E., Treyer, M., Arnouts, S., & Fouchez, D. 2019, A&A, 621, A26, doi: [10.1051/0004-6361/201833617](https://doi.org/10.1051/0004-6361/201833617)
- Pasquet-Itam, J., & Pasquet, J. 2018, A&A, 611, A97, doi: [10.1051/0004-6361/201731106](https://doi.org/10.1051/0004-6361/201731106)
- Polsterer, K. L., D'Isanto, A., & Gieseke, F. 2016, arXiv e-prints, arXiv:1608.08016, doi: [10.48550/arXiv.1608.08016](https://doi.org/10.48550/arXiv.1608.08016)
- Rau, M. M., Seitz, S., Brimoulle, F., et al. 2015, MNRAS, 452, 3710, doi: [10.1093/mnras/stv1567](https://doi.org/10.1093/mnras/stv1567)
- Rauch, M. 1998, ARA&A, 36, 267, doi: [10.1146/annurev.astro.36.1.267](https://doi.org/10.1146/annurev.astro.36.1.267)
- Richards, G. T., Weinstein, M. A., Schneider, D. P., et al. 2001, AJ, 122, 1151, doi: [10.1086/322132](https://doi.org/10.1086/322132)
- Sadeh, I., Abdalla, F. B., & Lahav, O. 2016, PASP, 128, 104502, doi: [10.1088/1538-3873/128/968/104502](https://doi.org/10.1088/1538-3873/128/968/104502)
- Schlafly, E. F., Meisner, A. M., & Green, G. M. 2019, ApJS, 240, 30, doi: [10.3847/1538-4365/aafbea](https://doi.org/10.3847/1538-4365/aafbea)
- Schlegel, D., Dey, A., Herrera, D., et al. 2021, in American Astronomical Society Meeting Abstracts, Vol. 237, American Astronomical Society Meeting Abstracts #237, 235.03
- Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv e-prints, arXiv:1503.03757, doi: [10.48550/arXiv.1503.03757](https://doi.org/10.48550/arXiv.1503.03757)
- Sutskever, I., Vinyals, O., & Le, Q. V. 2014, arXiv e-prints, arXiv:1409.3215, doi: [10.48550/arXiv.1409.3215](https://doi.org/10.48550/arXiv.1409.3215)
- Tabasi, S. S., Salmani, R. V., Khaliliyan, P., & Firouzjaee, J. T. 2023, ApJ, 954, 164, doi: [10.3847/1538-4357/ace03f](https://doi.org/10.3847/1538-4357/ace03f)
- Tanaka, M., Coupon, J., Hsieh, B.-C., et al. 2017, Publications of the Astronomical Society of Japan, 70, S9, doi: [10.1093/pasj/psx077](https://doi.org/10.1093/pasj/psx077)
- Teixeira, G., Bom, C. R., Santana-Silva, L., et al. 2024, Astronomy and Computing, 49, 100886, doi: [10.1016/j.ascom.2024.100886](https://doi.org/10.1016/j.ascom.2024.100886)
- Wadadekar, Y. 2005, PASP, 117, 79, doi: [10.1086/427710](https://doi.org/10.1086/427710)
- Way, M. J., & Klose, C. D. 2012, PASP, 124, 274, doi: [10.1086/664796](https://doi.org/10.1086/664796)
- Way, M. J., & Srivastava, A. N. 2006, ApJ, 647, 102, doi: [10.1086/505293](https://doi.org/10.1086/505293)
- Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2013, PhR, 530, 87, doi: [10.1016/j.physrep.2013.05.001](https://doi.org/10.1016/j.physrep.2013.05.001)
- Weinstein, M. A., Richards, G. T., Schneider, D. P., et al. 2004, ApJS, 155, 243, doi: [10.1086/425355](https://doi.org/10.1086/425355)
- Yao, L., Qiu, B., Luo, A. L., et al. 2023, MNRAS, 523, 5799, doi: [10.1093/mnras/stad1842](https://doi.org/10.1093/mnras/stad1842)
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579, doi: [10.1086/301513](https://doi.org/10.1086/301513)
- Zaremba, W., Sutskever, I., & Vinyals, O. 2014, arXiv e-prints, arXiv:1409.2329, doi: [10.48550/arXiv.1409.2329](https://doi.org/10.48550/arXiv.1409.2329)
- Zhang, C., Wang, W., Qu, M., Jiang, B., & Zhang, Y. 2024, AJ, 168, 244, doi: [10.3847/1538-3881/ad79f9](https://doi.org/10.3847/1538-3881/ad79f9)
- Zhou, X., Gong, Y., Meng, X.-M., et al. 2022, Research in Astronomy and Astrophysics, 22, 115017, doi: [10.1088/1674-4527/ac9578](https://doi.org/10.1088/1674-4527/ac9578)
- Zhou, X., Gong, Y., Meng, X.-M., et al. 2021, ApJ, 909, 53, doi: [10.3847/1538-4357/abda3e](https://doi.org/10.3847/1538-4357/abda3e)